E-Commerce Letters 电子商务评论, 2025, 14(4), 3020-3025
Published Online April 2025 in Hans. https://www.hanspub.org/journal/ecl
https://doi.org/10.12677/ecl.2025.1441219

Hans汉斯

人工智能时代电子商务平台内容审核的合规边界与法律挑战

丁子涵

扬州大学法学院，江苏 扬州

收稿日期：2025年3月13日；录用日期：2025年3月26日；发布日期：2025年4月30日

摘　要

随着数字经济飞速发展，人工智能技术在电子商务平台中的广泛应用推动了内容审核方式变革。自动化算法驱动的审核系统大幅提高了处理效率，但也带来了诸多法律合规挑战。本文探讨了人工智能时代电子商务平台内容审核的合规边界，重点分析了算法决策合法性、算法偏见与反歧视原则的冲突、透明度缺失与用户知情权的张力等法律问题。同时，探讨了跨境合规标准冲突的挑战及其可能的解决路径。为应对这些挑战，提出了构建算法全生命周期动态评估机制、重构平台责任认定、推行算法沙盒与穿透式监管等建议，希望对电子商务平台内容审核系统的合规性与公正性提供参考。

关键词

人工智能，电子商务，内容审核，法律合规

Compliance Boundaries and Legal Challenges of Content Review for E-Commerce Platforms in the Era of Artificial Intelligence

Zihan Ding

School of Law, Yangzhou University, Yangzhou Jiangsu

Received: Mar. 13th, 2025; accepted: Mar. 26th, 2025; published: Apr. 30th, 2025

Abstract

With the rapid development of the digital economy, the widespread application of artificial intelligence technology in e-commerce platforms has driven the transformation of content review meth-

文章引用: 丁子涵. 人工智能时代电子商务平台内容审核的合规边界与法律挑战[J]. 电子商务评论, 2025, 14(4): 3020-3025. DOI: 10.12677/ecl.2025.1441219

ods. The automated algorithm driven auditing system has significantly improved processing efficiency, but it has also brought many legal compliance challenges. This article explores the compliance boundaries of content review on e-commerce platforms in the era of artificial intelligence, with a focus on analyzing legal issues such as the legality of algorithm decisions, the conflict between algorithm bias and anti discrimination principles, the lack of transparency, and the tension between users' right to know. At the same time, the challenges of cross-border compliance standard conflicts and possible solutions were discussed. To address these challenges, suggestions have been put forward such as building a dynamic evaluation mechanism for the entire lifecycle of algorithms, restructuring platform responsibility determination, promoting algorithm sandboxing and penetrating supervision, hoping to provide reference for the compliance and fairness of e-commerce platform content review systems.

Keywords

Artificial Intelligence, Electronic Commerce, Content Review, Legal Cooperation

1. 引言

数字经济的蓬勃发展与人工智能技术的深度融合，正在重塑电子商务平台的治理生态。作为连接数亿用户与商家的核心枢纽，电子商务平台承载着商品交易、社交互动与信息传播的多重功能，其内容审核机制直接关乎市场秩序、用户权益与社会公共利益。在人工智能技术驱动下，平台内容审核从人工筛查转向算法主导的自动化模式，这种技术迭代虽显著提升了处理海量用户生成内容的效率，却也使得法律合规问题日益复杂化：技术逻辑与法律规则之间的张力，正在引发全球范围内的监管反思与制度重构。当前，人工智能审核系统的法律挑战集中体现于三重矛盾：算法效率与权利保护的冲突、技术黑箱与程序正义的悖论、全球合规与属地监管的错位。在此环境下，传统法律框架对“技术中立”原则的固守已难以适应人工智能审核系统的动态性特征。当算法通过实时反馈机制不断优化决策模型时，审核规则的实际运行可能偏离预设的合规路径，形成监管部门难以捕捉的合规暗区。本文旨在分析电子商务平台内容审核的合规边界，厘清审核机制与法律规制的交互管理关系，在此基础上提出解决对策，以期在技术创新与权利保障之间探寻动态平衡的法治路径。

2. 人工智能时代电子商务平台内容审核的合规边界

2.1. 算法决策的合法性基础与法律授权边界

人工智能技术在内容审核中的合法性基础需以现行法律框架为依托。根据《电子商务法》第 30 条，平台需对商品和服务信息建立审查机制，但未明确赋予算法自动化决策的授权范围。《网络安全法》《数据安全法》虽要求平台履行内容管理义务，但技术手段的适用边界仍存模糊。算法决策的合法性争议集中于“人工介入”的必要性，欧盟《数字服务法案》第 14 条要求对算法删除内容设置人工复核通道，而中国《网络信息内容生态治理规定》第 10 条强调“技术 + 人工”双重审核，但未规定人工介入的具体比例[1]。实践中，平台为提升效率往往将 99%的 UGC 交由算法处理，仅对高风险内容保留人工干预，这种技术优先的实践可能构成对法律精神的实质性架空。法律授权的核心矛盾在于，立法者预设的技术

工具论与算法实际形成的自主决策权之间存在结构性冲突，亟需通过立法明确算法审核的适用场景、决策层级及例外情形。

2.2. 算法偏见与反歧视原则的合规冲突

人工智能审核系统的训练数据隐含的社会偏见可能引发系统性歧视。例如，斯坦福大学 2023 年研究显示，针对特定方言、小众文化符号的误判率高达 35%，导致《反不正当竞争法》第 11 条禁止的“歧视性待遇”风险。欧盟《人工智能法案》(DSA)将内容审核系统列为高风险 AI，要求进行偏差测试和矫正，而中国《个人信息保护法》第 24 条要求自动化决策需保证结果公平[2]。但技术层面，数据偏差的修复面临困境，电商平台用户画像常基于地域、消费习惯等敏感信息，算法为提升审核精度可能强化群体特征识别，形成《消费者权益保护法》第 16 条禁止的“大数据杀熟”变体。合规难点在于，法律要求的结果公平与算法优化的特征关联存在根本性矛盾，需通过动态偏差监测、第三方审计等机制重构公平性标准。

2.3. 透明度缺失与用户知情权的法律张力

算法黑箱特性严重冲击《电子商务法》第 34 条规定的用户知情权。当商品链接因算法误判违规下架时，平台通常仅提供“违反社区规范”的格式化说明，用户无法获取具体违规标签、数据特征或决策路径信息。欧盟 DSA 第 15 条强制要求平台公开算法审核的主要参数与逻辑框架，但技术机密保护与公众知情权的平衡尚未解决。中国《互联网信息服务算法推荐管理规定》第 12 条要求公示算法基本原理，但电商平台多以“商业秘密”为由拒绝披露核心模型[3]。进一步地，算法的自我迭代机制使审核规则动态变化，即便初始版本合规，经过数月训练后可能衍生出法律未预见的新型过滤规则。这要求法律建立算法版本追溯制度，并赋予监管机关实时访问算法决策日志的权限，但此类措施可能遭遇企业以技术安全为由的抵制。

2.4. 跨境场景下合规标准的冲突与协调

全球化电商平台面临法域合规标准的分裂化挑战。例如，亚马逊需同时遵守欧盟 DSA 对“非法内容”的 24 小时删除令、美国《通信规范法》第 230 条的平台责任豁免规则，以及中国《网络安全法》对违法信息的立即阻断要求。不同法域对违法内容的定义差异显著，同一宗教服饰商品在欧洲可能因文化包容政策被允许展示，在东南亚国家可能触发宗教敏感审核。人工智能系统为适应多国合规要求，往往建立数百个区域性审核模型，导致合规成本指数级增长。更严峻的是，部分国家要求平台开放算法源代码供审查，如印度 2023 年《数字主权法案》草案，这与企业知识产权保护形成直接冲突[4]。解决路径可能在于国际组织推动审核标准的互认机制，例如借鉴 APEC 跨境隐私规则(CBPR)构建内容审核白名单，但意识形态与法律文化的差异使此类协调举步维艰。

3. 人工智能时代电子商务平台内容审核的法律挑战

3.1. 算法效率与权利保护的冲突

人工智能技术驱动的自动化审核系统虽极大提升了电子商务平台处理海量内容的能力，但其效率导向的技术逻辑与法律体系中的权利保护原则正面临深刻冲突。以深度学习为核心的算法模型通过数据训练形成决策规则，据 2023 年国际电商联盟报告，在识别违禁商品、虚假宣传等场景中展现出高达 90%的准确率，但效率提升的代价是用户权利的系统性压缩[5]。例如，算法对“医疗保健”“减肥”等关键词的过度敏感可能导致合规商品被误判下架，商家依据《电子商务法》第 35 条享有的公平交易权因此受损；基于用户行为数据的风险预测模型可能对特定消费群体实施差异化审核策略，触发《个人信息保护法》

第 24 条禁止的“自动化决策不公”问题。更严峻的是，算法效率与权利保护的冲突已嵌入技术架构底层，为提高审核速度，平台普遍采用“先处置后申诉”机制，根据中国消费者协会 2024 年调研数据，用户商品或账号被误封后需经历平均 7 个工作日的复核期，这直接违背《网络安全法》第 49 条“及时纠正错误处置”的法定要求[6]。法律应对的困境在于，现行立法对“审核效率”与“权利救济”的平衡缺乏量化标准，欧盟 DSA 试图通过“分层审核”制度缓解冲突，但实践中平台常通过降低风险评级阈值来规避义务，导致制度空转。

3.2. 技术黑箱与程序正义的悖论

人工智能审核系统的不可解释性正瓦解传统程序正义的制度基石，形成“算法暴政”与法治原则的对峙。深度学习模型的非线性决策特征使得即便平台运营者亦难以准确追溯内容下架的具体逻辑路径，这直接冲击《电子商务法》第 34 条规定的说明义务。当用户因“涉嫌售假”被强制下架商品时，平台仅能提供系统检测异常的模糊解释，商家既无法获取违规图像的特征向量数据，也不能质疑算法模型的置信度阈值设定，申诉权实质沦为形式。欧盟《通用数据保护条例》(GDPR)第 22 条赋予用户拒绝纯自动化决策的权利，但在电商场景中，用户若拒绝算法审核往往意味着服务终止，权利行使陷入“全有或全无”的困局。同时，技术黑箱还衍生出监管套利空间，某头部平台 2023 年被曝利用算法漏洞，对自营商品降低审核标准，使第三方竞品下架率高出 12 个百分点，该行为虽违反《反垄断法》第 17 条，但监管部门因无法解析算法决策规则而难以取证[7]。此外，程序正义的崩塌更体现在算法动态迭代对法律确定性的消解，一个通过合规测试的审核模型可能在三个月后因对抗训练产生歧视性规则，而现行法律缺乏对算法版本变更的动态监督机制。

3.3. 全球合规与属地监管的错位

电子商务平台的全球化运营正陷入监管割据困境，人工智能审核系统被迫在相互冲突的法域要求中寻求脆弱平衡。欧盟 DSA 要求平台对“非法内容”实施 24 小时内删除，并公开算法审核参数；美国《通信规范法》第 230 条则赋予平台对用户内容的广泛责任豁免；中国《网络信息内容生态治理规定》明确九类违法信息需实时阻断。例如，一款印有宗教符号的服饰在德国被算法识别为文化商品正常展示，在印尼因触发宗教敏感过滤机制自动下架，而平台可能同时面临欧盟市场监管局的“过度审查”指控与印尼通信部的“审查不力”处罚[8]。属地监管的刚性要求迫使企业部署区域性审核模型，据路透社 2024 年调查，头部电商平台需维护超过 200 个定制化算法模块，合规成本较五年前激增 470% [9]。更深层危机源于数据主权与算法透明度的矛盾，印度《数字主权法案》草案要求在本国运营的外国平台公开算法源代码，而中国《数据安全法》第 21 条禁止关键数据出境，这使跨境算法训练陷入合法性危机。国际协调机制的缺失加剧了合规困境，尽管经合组织(OECD) 2023 年发布《人工智能审核标准指南》，但成员国对“最小必要审核强度”“用户权利救济等级”等核心条款仍存在根本分歧。

4. 人工智能时代电子商务平台内容审核治理建议

4.1. 构建算法全生命周期动态影响评估体系

应对人工智能审核系统的合规风险，需建立覆盖算法设计、部署、迭代全流程的技术治理框架。在算法设计阶段，强制实施偏差检测与矫正机制，例如引入对抗性测试数据集模拟少数民族语言、小众文化符号等边缘场景，将《个人信息保护法》第 55 条要求的“数据影响评估”扩展为“算法公平性认证”。部署阶段推行可解释人工智能(XAI)技术嵌入，通过局部可解释模型或注意力热图将算法决策依据转化为可视化报告，满足《互联网信息服务算法推荐管理规定》第 12 条的信息披露义务[10]。在持续运营阶段，

建立算法版本追溯与决策日志存证系统，要求平台保留至少三年的算法变更记录及对应审核结果数据，为监管机构依据《电子商务法》第 33 条实施事后审计提供技术基础。针对算法动态迭代引发的合规漂移问题，可借鉴欧盟《人工智能法案》的“实时监测”条款，强制平台在模型更新后 72 小时内向监管部门提交影响评估报告，重点检测敏感群体误判率、规则偏离度等关键指标。技术治理的核心在于将法律原则转化为可量化的技术参数，例如将《反不正当竞争法》第 11 条“公平竞争”要求具象化为商家内容下架率差异不得超过基准值 2 个标准差，通过技术手段固化合规底线。

4.2. 重构技术性合规标准与责任分配规则

现行法律框架对人工智能审核的规制存在结构性缺陷，需通过立法技术革新构建适配算法特性的制度体系。首先，在《网络安全法》《电子商务法》中增设“算法审核特别条款”，明确自动化决策的适用边界：将商品信息真实性核验等低风险场景划归算法主导，而对知识产权侵权判定等高复杂度任务保留人工终审权。其次，重构平台责任认定规则，引入算法决策可抗辩推定原则，当用户证明算法存在系统性偏差，如特定类目商品误判率超过 15%时，平台需自证技术合理性，否则承担《消费者权益保护法》第 55 条的惩罚性赔偿责任。针对技术黑箱导致的举证难题，可建立“算法决策责任保险”制度，要求平台按审核模型风险等级缴纳保费，将部分合规成本转移至金融市场。最后，需突破传统技术中立原则，在《数据安全法》实施细则中创设“算法主动合规义务”，要求平台每季度对审核模型进行反歧视校准，并将校准数据纳入国家算法治理数据库。制度创新的关键在于打破“要么全自动、要么全人工”的二元规制模式，构建风险分级响应机制，例如对农产品直播带货实施算法宽容度提升 20%的差异化审核策略，实现法律刚性与技术弹性的有机统一[11]。

4.3. 建立算法沙盒与穿透式监管协同机制

传统监管手段难以应对人工智能审核系统的复杂性，需构建试验性监管联合数据穿透的新型治理范式。首先，在监管工具创新方面，推广算法监管沙盒制度，允许平台在限定区域测试新型审核模型，监管部门通过虚拟孪生技术同步构建镜像系统，实时监测算法决策与法律要求的偏差值，此机制可有效缓解《数据安全法》第 21 条数据出境限制与技术创新需求的矛盾。其次，监管能力升级方面，建立国家级算法审计平台，依据《互联网信息服务算法推荐管理规定》第 23 条，对头部电商平台实施 API 级数据对接，直接抓取审核日志、特征权重等核心参数，运用监管科技进行多维度合规分析。针对跨境监管难题，探索“等效审查”机制，对通过欧盟 DSA 合规认证的审核模块，可在中国市场免于重复检测，但需补充意识形态安全专项评估[12]。最后，在执法层面，赋予市场监管部门“算法紧急制动权”，当发现某审核模型引发大规模权利侵害时，可依据《反垄断法》第 48 条强制暂停算法运行，切换至人工审核模式。穿透式监管的核心是通过技术赋能使法律规则直达算法内核，例如将《网络信息内容生态治理规定》第 8 条“弘扬社会主义核心价值观”要求转化为审核模型的注意力权重参数，实现价值导向的技术化嵌入。

4.4. 推动跨境审核标准互认与治理共识形成

国际协调的终极目标是通过监管等效性实现多方共赢，例如承认中国“算法备案制”与欧盟“算法审计制”具有同等监管效力，但需在跨境电商场景中设置意识形态审查豁免条款，形成兼具包容性与约束力的新型数字治理秩序。破解全球合规困局需构建多层级的国际合作框架。首先，依托世界贸易组织(WTO)电子商务谈判机制，推动建立人工智能审核系统的国际认证体系，包括数据偏差容忍度、决策可解释性等级等核心指标，使通过认证的算法模块可在签署国间通用，降低重复合规成本。在法律协调层面，借鉴《联合国国际货物销售合同公约》(CISG)的契约精神，制定《跨境内容审核冲突法公约》，明确

算法审核引发的民事纠纷适用“最密切联系原则”，例如，当法国消费者因算法误判向中国平台索赔时，可优先适用欧盟 GDPR 的赔偿标准，但平台可援引中国《网络安全法》证明技术合理性[13]。其次，发起“人工智能审核全球对话机制”，设立宗教敏感、文化符号等特定内容的“负面清单缓冲带”，例如对一些特殊商品设置 72 小时人工复核期，避免算法直接触发下架。针对数据主权争议，推广“算法联邦学习”模式，例如海外商家部署的审核模型通过加密参数聚合进行本地化训练，既满足《数字主权法案》的数据本地化要求，又保持核心算法架构的全球统一性。

5. 结论

综上所述，在人工智能驱动的电子商务平台内容审核系统不断发展的背景下，法律框架亟需与技术进步同步调整。本文通过分析算法决策合法性、偏见问题、透明度缺失及跨境监管等法律挑战，提出了相应的法律治理建议。虽然当前法律体系存在许多挑战，但通过立法创新、跨境合作与技术治理的推进，有望建立一个既能推动技术进步，又能保障用户权利的合规环境。未来，随着法律和技术的不断协调与优化，电子商务平台的内容审核机制将更加健全，为全球数字经济的健康发展提供有力保障。

参考文献

[1] 李泳佳. 数字经济背景下跨境电子商务发展的机遇与挑战研究[J]. 中国商论, 2025, 34(3): 39-42.

[2] 王诤毅. 电子商务平台经营者侵权责任的多维度探析[J]. 法制博览, 2025(3): 64-66.

[3] 陈旭, 魏冰. 跨模态内容监测与审核应用探讨[J]. 广播电视信息, 2025, 32(1): 97-100.

[4] 林洁, 王莉萍. 部分国家主要电子商务法规对比分析[J]. 法制博览, 2025(1): 10-12.

[5] 张惠彬, 张秀璇. 公私法协同视野下 AIGC 服务提供者的内容审核义务[J]. 武汉科技大学学报(社会科学版), 2024, 26(6): 62-73.

[6] 刘美. 论电子商务经营者的信息义务——对《电子商务法》第 50 条的评注[J]. 湖南科技大学学报(社会科学版), 2024, 27(5): 119-126.

[7] 高扩, 李钢. 平台型媒体内容审核与风险防范策略探析[J]. 中国报业, 2024(15): 17-19.

[8] 买迪努尔·买买提.《中华人民共和国电子商务法》的实施与制度的完善发展[J]. 现代商贸工业, 2024, 45(14): 182-184.

[9] 尚琦. 完善《电子商务法》中关于“通知-删除”规则的必要性分析[J]. 法制博览, 2024(12): 120-122.

[10] 包志法, 王春娟, 顾奕, 等. “三位一体”重大决策合法合规性审核体系的创新实践[J]. 企业家, 2023(S1): 108-111.

[11] 燕鹏琦. 电子商务平台经营者安全保障义务研究[D]: [硕士学位论文]. 天津: 天津商业大学, 2023.

[12] 徐国瀚. 论电子商务平台经营者违反审核义务的侵权责任[D]: [硕士学位论文]. 济南: 山东大学, 2023.

[13] 卢海波, 骆迅, 唐晔, 等. AI 赋能 + 合规导向 + 系统闭环: 芒果 TV 内容安全保障技术体系的构建[J]. 广播电视信息, 2022, 29(6): 21-24.